# FLOATING GATE MEMORY STRUCTURES
# AND FABRICATION METHODS

## Chia-Shun Hsiao

5 ## BACKGROUND OF THE INVENTION

The present invention relates to floating gate nonvolatile memories.

A floating gate nonvolatile memory cell stores information by storing an electrical charge on its floating gate. The floating gate is capacitively coupled to the control gate. In order to write the cell, a potential difference is created between the control gate and
10 some other region, for example, the source, drain or channel region of the cell. The voltage on the control gate is capacitively coupled to the floating gate, so a potential difference appears between the floating gate and the source, drain or channel region. This potential difference is used to change the charge on the floating gate.

In order to reduce the potential difference that has to be provided between the
15 control gate and the source, drain or channel region, it is desirable to increase the capacitance between the control and floating gates relative to the capacitance between the floating gate and the source, drain or channel region. More particularly, it is desirable to increase the "gate coupling ratio" GCR defined as $C_{CG}/(C_{CG} + C_{SDC})$ where $C_{CG}$ is the capacitance between the control and floating gates and $C_{SDC}$ is the capacitance between
20 the floating gate and the source, drain or channel region. One method for increasing this ratio is to form spacers on the floating gate. See U.S. patent no. 6,200,856 issued March 13, 2001 to Chen, entitled "Method of Fabricating Self-Aligned Stacked Gate Flash Memory Cell". In that patent, the memory is fabricated as follows. Silicon substrate 104 (Fig. 1) is oxidized to form a pad oxide layer 110. Silicon nitride 120 is formed on oxide
25 110 and patterned to define isolation trenches 130. Oxide 110 and substrate 104 are etched, and the trenches are formed. Dielectric 210 (Fig. 2), for example, borophosphosilicate glass, is deposited over the structure to fill the trenches, and is planarized by chemical mechanical polishing (CMP). The top surface of dielectric 210 becomes even with the top surface of nitride 120. Then nitride 120 is removed (Fig. 3).
30 Oxide 110 is also removed, and gate oxide 310 is thermally grown on substrate 104 between the isolation trenches. Doped polysilicon layer 410.1 (Fig. 4) is deposited over

the structure to fill the recessed areas between the isolation regions 210. Layer 410.1 is polished by chemical mechanical polishing so that the top surface of layer 410.1 becomes even with the top surface of dielectric 210.

5    Dielectric 210 is etched to partially expose the edges of polysilicon layer 410.1 (Fig. 5). Then doped polysilicon 410.2 is deposited and etched anisotropically to form spacers (Fig. 6) on the edges of polysilicon 410.1. Layers 410.1, 410.2 provide the floating gates.

As shown in Fig. 7, dielectric 710 (oxide/nitride/oxide) is formed on polysilicon 410.1, 410.2. Doped polysilicon layer 720 is deposited on dielectric 710 and patterned to
10    provide the control gates.

Spacers 410.2 increase the capacitance between the floating and control gates by more than the capacitance between the floating gates and substrate 104, so the gate coupling ratio is increased.


SUMMARY

15    This section is a brief summary of some features of the invention. The invention is defined by the appended claims which are incorporated into this section by reference.

In some embodiments of the present invention, the gate coupling ratio is increased by making the trench dielectric regions 210 more narrow at the top (see Fig. 14 for example). Therefore, the floating gate polysilicon layer is wider at the top (see Fig. 15).
20    This increased width improves the gate coupling ratio. A single polysilicon layer is sufficient to form the floating gates with the increased gate coupling ration, though multiple polysilicon layers can also be used.

Other features are described below.


BRIEF DESCRIPTION OF THE DRAWINGS

25    Figs. 1-7 show cross sections of prior art nonvolatile memory structures in the process of fabrication.

Figs. 8-16 show cross sections of nonvolatile memory structures in the process of fabrication according to the present invention.

Fig. 17 is a circuit diagram of a memory array according to the present invention.

Fig. 18 is a top view of the memory of Fig. 17.

Figs. 19A, 19B show cross sections of the memory of Fig. 17.

The following table describes some reference numerals used in the drawings.

104 ........................... substrate
110 ......................... pad oxide
120 ................... silicon nitride
130 ............. isolation trenches
210 ............... trench dielectric
310 ......................... gate oxide
410, 410.1, 410.2 ........... floating gate layers
710 ......................... dielectric
720 ..................... control gates
810 ................. silicon dioxide
814 ................... silicon nitride
820 ..................... photoresist
1720 ...................... wordlines
1820 ......... source line regions
1830 ................ silicon nitride
1840 .............. stack structures
1850 ...................... dielectric

DESCRIPTION OF PREFERRED EMBODIMENTS

This section describes some embodiments to illustrate the invention. The invention is not limited to these embodiments. The materials, conductivity types, layer thicknesses and other dimensions, circuit diagrams, and other details are given for illustration and are not limiting.

Fig. 8 illustrates the beginning stages of fabrication of a memory array according to one embodiment of the invention. An isolated doped region of type P- is formed in monocrystalline semiconductor substrate 104 as described, for example, in U.S. patent no. 6,355,524 issued March 12, 2002 to H.T. Tuan et al. and incorporated herein by reference. This region is isolated by P-N junctions (not shown). Other isolation techniques, and non-isolated regions, can also be used.

Silicon dioxide layer 110 (pad oxide) is formed on substrate 104 by thermal oxidation or some other technique to an exemplary thickness of 9 nm. Silicon nitride 120 is deposited on oxide 110. An exemplary thickness of this layer is 90 nm. Another

silicon dioxide layer 810 is formed on nitride 120. An exemplary thickness of this layer is 5 nm. Silicon nitride 814 is deposited on oxide 810, to a thickness of 90 nm.

Photoresist mask 820 is formed on layer 814 by means of photolithography. This mask defines (and exposes) isolation trenches 130 (Fig. 9). This mask also defines (and covers) substrate areas 132 not occupied by the isolation trenches. Areas 132 include the active areas (the source, drain and channel regions) of the memory cells.

Layers 814, 810, 120, 110, and substrate 104 are etched where exposed by the mask, to form the isolation trenches. (Resist 820 can be removed immediately after the etch of nitride 814 or at a later stage.)

Then dielectric 210 (Fig. 10) is formed to fill the isolation trenches and cover the structure. Dielectric 210 can include as a combination of layers including a thick final layer of silicon dioxide deposited by chemical vapor deposition (CVD) using high density plasma. See the aforementioned U.S. patent no. 6,355,524.

Dielectric 210 is polished by CMP until nitride 814 is exposed. The top surface of dielectric 210 is about even with the top surface of nitride 814.

Nitride 814 is removed selectively to dielectric 210 (Fig. 11). This can be done by a wet etch (e.g. with phosphoric acid).

Then dielectric 210 is etched (Fig. 12). This etch includes a horizontal component that causes the sidewalls of dielectric 210 to be laterally recessed away from areas 132. This etch can also remove the oxide 810. The etch can be an isotropic wet etch selective to silicon nitride. A buffered oxide etch or a dilute HF (DHF) etch is used in some embodiments.

The resulting profile of dielectric 210 is a function of the etch process and the thicknesses and composition of layers 110, 120, 810, 814. Fig. 13 shows the top portion of dielectric 210 on a larger scale. The dotted line at the top marks the shape of dielectric 210 before the etch. Dimension "y" is the amount by which the dielectric 210 is etched vertically. Dimension "x" is the amount by which the sidewall is recessed horizontally at the top. Dimension "z" is the amount by which the bottom edge of the recessed sidewall portion is below the top surface of dielectric 210 at the end of the etch. The wet etch described above is isotropic, so $x = y = z$. The amount by which the bottom edge of the recessed sidewall is below the surface of nitride 120 is a function of the thickness of oxide 810. This amount is also a function of the etch selectivity relative to silicon nitride.

- 4 -

The selectivity is practically infinity in some embodiments. The profile of the resulting structure is also affected by the thickness of layers 110, 120 and the etch duration. Different profiles of dielectric 210 can thus be obtained. In Fig. 13, the dielectric sidewalls curve laterally away from areas 132 as the sidewalls are traced upward.

5      Silicon nitride 120 and oxide 110 are removed (see Fig. 14). The etch of oxide 110 also removes a portion of oxide 210. This is an anisotropic etch in some embodiments.

Turning now to Fig. 15, silicon dioxide 310 (tunnel oxide) is thermally grown on the exposed areas 132 of substrate 104. An exemplary thickness of oxide 310 is 9 nm.

10      Polysilicon layer 410 (floating gate polysilicon) is formed to fill the areas between dielectric regions 210 and cover the structure. Polysilicon 410 is polished by CMP until the dielectric 210 is exposed. Layer 410 is made conductive by doping. The horizontal top surface of polysilicon 410 projects over the isolation trenches 130 laterally beyond the areas 132.

15      Floating gates 410 abut dielectric regions 210. In Fig. 15, the floating gate sidewalls extend laterally outward beyond areas 132 as the sidewalls are traced upward. Different sidewall profiles can be obtained as defined by the sidewall profiles of dielectric 210.

Then ONO 710 (Fig. 16) is formed over the structure, and control gate polysilicon 20      720 is deposited and patterned. Polysilicon 720 is made conductive by doping. Layers 710, 410 can be patterned after the patterning of layer 720 as appropriate.

A wide range of floating gate memories can be made using the teachings of the present invention, including stacked gate, split gate and other cell structures, flash and non-flash EEPROMs, and other memory types known or to be invented. An example 25      split gate flash memory array is illustrated in Figs. 17, 18, 19A, 19B. This memory array is similar to one disclosed in the aforementioned U.S. patent no. 6,355,524 but is modified to increase the gate coupling ratio. Fig. 17 is a circuit diagram of the array. Fig. 18 is a top view. Fig. 19A is a cross section along the line A-A in Fig. 18. Line A-A passed through a control gate line 720 providing the control gates for one row of the 30      memory cells. Fig. 19B is a cross section along the line B-B which passes through a bitline 1704 extending across the array in the column direction.

Each memory cell 1710 includes a floating gate 410, a control gate 720, and a select gate 1720. The control gates lines 720 are made of doped polysilicon. The select gates for each row are provided by a doped polysilicon wordline. Wordlines 1720 and control gate lines 720 extend in the row direction across the array. In Fig. 17, each

5 memory cell is shown schematically as a floating gate transistor and an NMOS transistor connected in parallel.

Each memory cell has source/drain regions 1810, 1820. Regions 1810 ("bitline regions") are adjacent to the select gates. These regions are connected to the bitlines. Regions 1820 ("source line regions") of each row are shared with regions 1820 of an

10 adjacent row on the opposite side of the cells from regions 1810. Regions 1820 of the two rows are merged into a diffused source line that runs in the row direction across the array.

Isolation trenches 130 are placed between adjacent columns of the array. The trench boundaries are shown at 130B in Fig. 18. Each trench runs under two adjacent

15 rows of the array (under two control gate lines 720 and respective wordlines 1720) and terminates at source lines 1820, slightly projecting into the source lines from under the control gate lines. Floating gates 410 overlap the isolation trenches, as in Fig. 15.

Trenches 130, trench dielectric 210, tunnel oxide 310, floating gate layer 410, and dielectric 710 are manufactured as described above in connection with Figs. 8-16. Then

20 polysilicon 720 is deposited as described above. Silicon nitride 1830 is deposited over polysilicon 720 and patterned photolithographically to define the control gate lines 720. Layers 720, 710, 410, 310 are etched away in the areas not covered by nitride 1830. The remaining portions of nitride 1830, polysilicon 720, ONO 710, polysilicon 410, and oxide 310 form a number of stacks 1840. Each stack corresponds to one row of the array.

25 The remaining fabrication steps can be as in the aforementioned U.S. patent no. 6,355,524. Dielectric 1850 (Fig. 19B) is formed on the sidewalls of each stack to insulate the floating and control gates from the wordlines. Silicon dioxide 1860 is grown on the exposed portions of substrate 104 to provide gate dielectric for the select gates. Polysilicon 1720 is deposited and etched anisotropically without a mask over the array to

30 form spacers on the stack sidewalls. Then a masked etch of polysilicon 1720 removes those spacers that are not used for the wordlines (the spacers over the source line regions 1820). The same mask (not shown) can be used to dope the source lines 1820. Then the

mask is removed, and additional dopant is implanted to dope the source line and bitline regions 1810, 1820.

The invention is not limited to the embodiments described above. For example, pad oxide 110 (Fig. 8) can be omitted, or used as tunnel oxide 310 (Fig. 14). Oxide 810

5    can also be omitted; silicon nitride layers 120, 814 can be combined into a single layer. This layer can be etched at the stage of Fig. 11 with a timed etch. Alternatively, this layer can be completely removed before the etch of dielectric 210. The entire sidewall portion of dielectric 210 above substrate 104 can be laterally recessed by the etch. The invention is not limited to any particular materials or memory layouts or circuit diagrams. The

10   invention is defined by the appended claims.